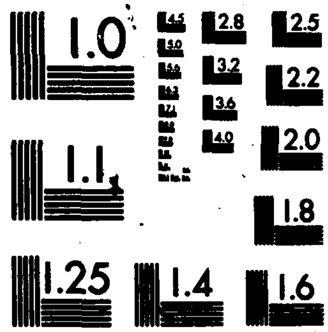


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

Bias and Information of Bayesian Adaptive Testing

David J. Weiss
James R. McBride

AD A129280

RESEARCH REPORT 83-2
MARCH 1983

COMPUTERIZED ADAPTIVE TESTING LABORATORY
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Army Research Institute, Air Force Office of Scientific Research, Air Force Human Resources Laboratory, and Office of Naval Research, and monitored by the Office of Naval Research

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
Research Report 83-2

2. GOVT ACCESSION NO.
AD-A129 980

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Bias and Information of Bayesian Adaptive Testing

5. TYPE OF REPORT & PERIOD COVERED
Technical Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
David J. Weiss and James R. McBride

8. CONTRACT OR GRANT NUMBER(s)
N00014-79-C-0172

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Psychology
University of Minnesota
Minneapolis, Minnesota 55455

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
P.E.: 61153N Proj: RR042-04
T.A.: RR042-04-01
W.U.: NR 150-433

11. CONTROLLING OFFICE NAME AND ADDRESS
Personnel and Training Research Programs
Office of Naval Research
Arlington, Virginia 22217

12. REPORT DATE
March 1983

13. NUMBER OF PAGES
20

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This research was supported by funds from the Army Research Institute, the Air Force Office of Scientific Research, the Air Force Human Resources Laboratory, and the Office of Naval Research, and monitored by the Office of Naval Research.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Adaptive Testing; Tailored Testing; Computerized Testing; Ability Testing; Item Response Theory; Latent Trait Test Theory; Item Characteristic Curve Theory; Bayesian Testing; Bayesian Scoring; Test Information; Bias of Ability Estimates; Monte Carlo Simulation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Monte carlo simulation was used to investigate score bias and information characteristics of Owen's Bayesian adaptive testing strategy, and to examine possible causes of score bias. Factors investigated in three related studies included effects of item discrimination, effects of fixed vs. variable test length, and effects of an accurate prior θ estimate. Data were generated from a three-parameter logistic model for 3,100 simulees in each of eight data sets; Bayesian adaptive tests were administered, drawing items from a "per-

fect" item pool. Results showed that the Bayesian adaptive test resulted in unbiased θ estimates and relatively flat information functions only in the unrealistic situation in which an accurate prior θ estimate was used. When a more realistic constant prior θ estimate was used with a fixed test length, severe bias was observed, with low θ levels overestimated and high θ levels underestimated; bias decreased for high θ levels with increased item discrimination, but discrimination did not substantially affect bias for low θ levels. Information curves for the constant prior and fixed test length condition became more peaked and asymmetric with increasing item discrimination. A different pattern of bias was observed with variable test length and a constant prior. In this case, increasing discriminations resulted in higher levels of bias for low θ levels and lower levels of bias for high θ levels. Low discriminations resulted in a flatter information function, with equiprecise measurement decreasing with increasing item discrimination. Also in the variable test length condition the test length required to achieve a specified level of the posterior variance of θ estimates was an increasing function of θ level, with twice the number of items required at high θ levels than at low θ levels. These results indicate that θ estimates from Owen's Bayesian adaptive testing method are affected by the prior θ estimate used and that the method does not provide measurements that are unbiased and equiprecise except under the unrealistic condition of an accurate prior θ estimate.

Contents

ACKNOWLEDGMENTS

The assistance of Joel M. Brown in the analysis of these data is appreciated.

Technical Editor: Barbara Leslie Camm

BIAS AND INFORMATION OF BAYESIAN ADAPTIVE TESTING

Since test scores are typically used to differentiate among persons, one highly desirable property of a test would be that it measure equally well at all points. Another consideration is that it measure each person precisely. Thus, an "ideal" test would have a high, horizontal information function. Unfortunately, this ideal cannot normally be achieved in a fixed-length conventional test that draws its items from a much larger fixed pool of test items. Ordinarily, some trade offs must be made. Relatively high information at a point can be achieved by "peaking" the test, that is, constructing it of the most discriminating items in a narrow range of difficulty. A relatively flat but low information function can be achieved by selecting equidiscriminating items having a wide range of item difficulty values. The only way to approximate a high, flat information function is to administer to each person the subset of items that provides the most information at his/her level of ability, θ. The problem with this is obvious: θ is unknown before the test is administered.

An adaptive test can select items during the course of testing in such a way as to attempt to maximize the information obtained for each examinee. This may be done either by simple branching—administering a more difficult item after a correct answer and an easier item after an incorrect answer—or by more elaborate techniques. Owen's (1969, 1975) Bayesian adaptive testing strategy estimates θ after each item response, then selects the unused test item that is, in one sense, the most "informative" at the current estimated ability level. The result is that different persons take different sets of test items; each set of test items spans a range of difficulty levels approximately tailored to provide maximal information about the individual examinee.

The information function of the test scores derived from any adaptive testing procedure should be (1) flatter than that of a peaked test of the same length and constructed from the same item pool and (2) higher than that of a rectangular test of the same length drawn from the same item pool. The height of the adaptive test's information function will be determined in large part by the discriminations and guessing parameters of the constituent items of the item pool as well as by test length. The flatness of the information curve (and to some extent its height) will depend largely on the range of item difficulties in the pool and on the effectiveness of the adaptive item selection procedure.

Urry (1971) conducted monte carlo simulations of Owen's (1969, 1975) sequential procedure using three different simulated item banks: two banks of "ideal" item parameters and one bank of items with the same parameters as the VSAT (Lord, 1968). Urry's item Bank A had 20 equidiscriminating items ($\underline{a}$ = 1.6) at each of five equally spaced levels on the ability continuum; his Item Bank B employed five items of the same ($\underline{a}$ = 1.6) discriminations at each of 20 ability levels; and Item Bank C employed the parameters actually occurring in the VSAT. Banks A and B required an average of just over 11 items to test termination. Bank C required an average of 27.5 items to termination. The other noteworthy result of Urry's (1971) simulation studies was the magnitude of the fidelity coefficients. For simulated examinees drawn randomly from a normal (0,1) population, the observed correlations of .936 (Item Bank A) and .919 (Item Bank B) are quite high in view of the relatively short test lengths involved.

Jensema (1972) simulated Owen's (1969, 1975) approach to Bayesian testing using the actual item responses of 100 live examinees to 58 mathematics items drawn from four conventional pre-college tests taken at full length by the examinees. From a record of their item-by-item actual test performance, a computer program constructed artificial protocols of their responses to the items that would have been administered by Bayesian sequential tests under two different conditions: with and without differential prior information about examinees' abilities. Parallel to these two "real data" simulations, Jensema carried out monte carlo simulations of the Bayesian procedure. These simulations used 100 simulated examinees and items with logistic ogive parameters identical to the 58 real items. Item scores were generated as a stochastic function of ability, θ , and the parameters of each item. The adaptive tests were terminated in each instance when the posterior variance of the Bayesian ability estimate fell below .0625 or when 30 items had been administered, whichever occurred first.

In the real-data simulation, mean test length was about 27 items, with or without differential initial ability estimates. The Bayesian estimates correlated about .86 with scores on a weighted composite of the four conventional tests from which the item bank was selected. Jensema did not report a correlation of ability with test length or with precision of estimate, but he did observe that the posterior variance criterion terminated the testing only in the upper portions of the distribution of estimated ability. Jensema interpreted these results to imply that the item pool was unsatisfactory for adaptive testing in the lower ability levels due to the low discriminations of the items in that region of the difficulty continuum. His monte carlo results using the same item pool resulted in virtually identical mean test lengths and in correlations of .92 between estimated ability and true ability. He concluded, in part, that a satisfactory item pool for adaptive testing needs to employ very highly discriminating items uniformly distributed on the difficulty continuum. Another conclusion he reached--this one on the basis of monte carlo simulation with ideal item banks--was that for most purposes little was to be gained by the use of prior information about examinees to determine a variable initial θ estimate. Jensema found that using differential prior information resulted in an average savings of only one test item.

In another monte carlo study of Owen's Bayesian strategy, Jensema (1974) examined the effects of item parameters and Bayesian test length on test reliability. He showed that reliability is directly related to the posterior variance of the Bayesian ability estimate; hence, using a specific value of that posterior variance as a termination criterion determines the reliability of the test. Jensema showed that the average number of items required to attain that reliability varies as a function of the item parameters. With items uniformly distributed on difficulty, the higher the item discrimination, the shorter the test.

McBride (1977; McBride & Weiss, 1976) also studied characteristics of the ability estimates resulting from Owen's (1969, 1975) strategy. These monte carlo simulations involved (1) an ideal item pool with variable test length; (2) the effects of guessing and item discrimination in a perfect item pool; (3) the effects of fixed test length; and (4) the effects of ability level and item pool configuration. In the first three studies, the performance of the adaptive test was evaluated on overall indices including the overall bias and mean absolute error of the ability estimates, the correlation of ability estimates with true ability estimates (fidelity), and correlations of true and estimated ability levels with errors and test length.

The fourth study evaluated the performance of this testing strategy in an item pool with no correlation between difficulty and discrimination parameters, and using items with high negative and high positive correlations between these parameters. In contrast to the other studies, characteristics of the ability estimates were examined as a function of true θ; dependent variables included bias and information conditional on θ. Contrasting with the first three studies, which showed little overall mean bias and information, Study 4 showed severe bias in the conditional θ estimates for all three item pool configurations. Estimates of θ were unbiased only for five θ values between $\theta = 1.0$ to -1.0; for low θ values, θ was overestimated and high θ values were underestimated. In addition, the information curves for the three item pool configurations were not high and flat as would be expected, at least when the ideal item pool was used in which difficulty and discrimination parameters were uncorrelated.

Gorman (1980) also examined the bias and information of scores produced by Owen's Bayesian testing procedure. These analyses were based on two "ideal" item pools with discriminations of $\underline{a} = .8$ and 1.6, in which 101 items were rectangularly distributed in difficulty, and both true and estimated item parameters were used. Gorman also studied the effect of applying a correction for regression (proposed by Urry, 1977) to ability estimates from Owen's testing procedure, designed to reduce bias in the estimates. His results show substantial bias in the uncorrected θ estimates, with positive bias for θ levels below zero, negative bias for θ levels above zero, and higher levels of bias for the less discriminating items. His data also show that Urry's correction was not entirely successful in eliminating the bias, since the corrected θ estimates for θ levels above zero resulted in positive bias. Since Gorman's study used an ideal, but finite, item pool, however, his results may be partially item pool dependent. In addition, Gorman's study did not attempt to determine the cause of the bias in the θ estimates but simply examined one possible approach to reducing it.

Purpose

The present study was designed to further investigate the nature of the bias and the information characteristics of Owen's Bayesian adaptive testing strategy and to examine possible causes of the bias. Factors investigated included (1) the effects of item discrimination, (2) the effects of fixed vs. variable test length, and (3) the effect of an accurate prior θ estimate.

theta

Method

Design

Monte carlo simulation of Owen's adaptive test was used. Unlike some previous simulation studies, but similar to Studies 1 to 3 in McBride (1977), the present studies did not use a prestructured item pool. Rather, the tests were simulated using a perfect and infinite item pool having any difficulty parameters required by the item selection process, with restrictions only on the item discriminations and pseudo-guessing parameters, c. By thus simulating an infinite item pool, the results of the simulation studies should reveal, within the limits of sampling error, the inherent properties of the Bayesian adaptive test, unafffected by the idiosyncrasies of a typical finite item pool.

Similarly, following the procedures of Study 4 in McBride (1977) in order to permit accurate description of the properties of the testing method as they vary with trait level, the simulated examinees (simulees) were not drawn randomly from a specified distribution; rather, a large number of examinees were simulated at each of a number of trait levels throughout the normally encountered range.

Examinees

For the purposes of monte carlo simulation, an examinee i was characterized by a numerical value, which is the actual trait level θ. In each of the eight data sets generated, there were 3,100 simulees, with 100 at each of 31 θ levels equally spaced in the interval -3.0 to 3.0. This range of the trait would include 99.99% of a population normally distributed on θ, with mean 0 and variance 1.

Test Items

For each separate item administration, an item was computer generated with the pseudo-guessing (c) parameter held constant at .20, simulating a five-alternative multiple-choice item. The item discrimination, a, was constant for each data set, with a = .80, 1.60, or 2.40 between data sets.

Following McBride (1977) the difficulty (b) parameter for each simulated item administration was determined by the current θ (the prior mean M_{m-1} of the estimated distribution of θ_i before administering the mth item) and by the constant item parameters a_g and b_g, according to the formula

$$b_g = M_{m-1} - \frac{1}{1.7a_g} \log\left[\frac{1 + (1 + 8c_g)^{1/2}}{2}\right] \qquad [1]$$

Equation 1 gives the item difficulty value having maximal information when $\theta_i = M_{m-1}$, and a_g and c_g are fixed (Birnbaum, 1968, p. 464). Since, in general, θ_i is unknown and the best available estimate is M_{m-1}, the item difficulty chosen is the one that is the most informative, given the current estimate of θ at any point in the adaptive test.

Item Responses

The dichotomous (0,1) score of any simulee on any item is a probabilistic function of its status θ_i on the trait θ, the item difficulty b_g, and the parameters a_g and c_g. The probability $P'_g(\theta_i)$ of a correct response ($u_g = 1$) under the logistic model item characteristic curve is

$$P'_g(\theta_i) = c_g + (1-c_g)/\left\{1 + \exp\left[-1.7a_g(\theta_i-b_g)\right]\right\}. \qquad [2]$$

In order to simulate item responses, each time an item administration took place the quantity $P_g'(\theta_i)$ was compared with a pseudo-random number r_{gi} generated from a distribution uniform in the interval [0,1]. A score of $u_g = 1$ was assigned whenever $P_g'(\theta_i)$ equaled or exceeded r_{gi}; otherwise, a score of 0 was assigned.

Dependent Variables

For the simulated test of each individual $\underline{i}$, the following were recorded:
$\underline{k}$, the number of items administered;
M_k, the posterior mean after $\underline{k}$ items (i.e., $\hat{\theta}$); and
V_k, the posterior variance after $\underline{k}$ items (i.e., the variance of $\hat{\theta}$).
These values were averaged at each level of θ across the 100 simulees at that level, resulting in $\bar{\hat{\theta}}_i$, the mean of the θ estimates at each level of $\theta_i (i = 1, 2, ..., 31)$, and $\sigma^2(\theta_i)$, the variance of $\hat{\theta}$ at each θ level. Bias was determined at each of the θ levels by

$$\text{Bias} = (\bar{\hat{\theta}}_i - \theta_i) \qquad [3]$$

Information was computed from the formula

$$I(\theta_i) = \bar{\hat{\theta}}_i'^2 \Big/ \sigma^2(\hat{\theta}_i) \qquad [4]$$

where $\hat{\theta}_i'$ is the first derivate of the polynomial regression of $\hat{\theta}$ on θ.

Independent Variables

Eight data sets were analyzed for three levels of item discrimination. The characteristics of the three studies and the data sets are summarized in Table 1.

Study I: Accurate prior θ estimate. This study was intended to provide "best case" data in order to serve as a benchmark against which other studies could be evaluated. The "best case" for the Bayesian adaptive test ought to be one involving a "perfect" item pool and accurate prior knowledge about examinees' trait levels. Accurate prior knowledge means that each examinee's trait level was known beforehand and was used as the mean of the Bayes prior distribution. Under these conditions the only limitations on the information and accuracy of estimate of Owen's procedure are those imposed by the test length, and by the discriminations and guessing parameters of the simulated test items. Holding those variables constant, any idiosyncrasies in the behavior of the test scores must be due to the trait level estimation and item difficulty selection procedure.

Two separate and independent test administrations were simulated for each of the 3,100 simulees: in Data Set 1, all item discriminations were .80, and in Data Set 2, $\underline{a} = 1.60$. For each simulee, the Bayes initial prior distribution

Table 1
Summary of the Independent Variables
in the Three Studies

Study and Data Set	a	Prior Distribution Mean	Prior Distribution Variance	Termination Criterion Posterior Variance	Termination Criterion No. of Items
Study I					
1	.80	θ_i	1	-	20
2	1.60	θ_i	1	-	20
Study II					
3	.80	0	1	-	20
4	1.60	0	1	-	20
5	2.40	0	1	-	20
Study III					
6	.80	0	1	.10	30
7	1.60	0	1	.10	30
8	2.40	0	1	.10	30

was normal, with mean θ_i and variance 1.0. Thus, at the outset of testing, the initial estimate of each simulee's trait level was accurate. The adaptive test was allowed to run its normal course, re-estimating θ_i after every item response and selecting the next item accordingly, until 20 items had been administered.

<u>Study II: Constant prior θ estimate with fixed test length</u>. Study II replicated the 20-item fixed test length and constant *a* values of .80 and 1.60 from Study 1; to examinee effects with more highly discriminating items, Data Set 5 used *a* = 2.40 for all items, while Data Sets 3 and 4 used items with *a* = .80 and 1.60 as in Study I. In contrast to Study I, the three data sets of Study II used the same initial normal prior distribution (mean = 0, variance = 1.0) for all simulees, regardless of actual trait level. In this study, then, a more typical use of the Bayesian adaptive testing strategy was simulated, i.e., the application to individuals for whom no prior θ estimates were available prior to testing; consequently, a group prior θ distribution was used to select the first item to be administered. As in Study I, a fixed-length test of 20 items was administered to each simulee.

<u>Study III: Constant prior θ estimate with variable test length</u>. In Study III, as in Study II, the same initial normal (0,1) prior distribution was assumed for all simulees. The difference between the studies was in the test termination criterion. In Study III, testing was terminated for each simulee whenever the posterior variance V_k fell below .10. This value corresponds to the "standard error of estimate" criterion of .3162 specified by Urry (1974) to achieve a fidelity coefficient exceeding .95 in a normal (0,1) population of examinees. A maximum test length of 30 items was imposed, so that if the posterior variance criterion had not been reached within 30 items, testing was terminated. As for Study II, three levels of item discrimination—*a* = .80, 1.60, and 2.40—were studied in Data Sets 6, 7, and 8, respectively.

Results

Accurate Prior θ Estimate

Bias of the ability estimates for the two data sets of Study I are shown in Figure 1 (numerical values of bias and information for Data Sets 1 and 2 are in Appendix Table A). As Figure 1 shows, there was virtually no bias in the ability estimates for Data Set 2 (a = 1.6), with a small amount of bias alternating between positive bias and negative bias for Data Set 1 (a = .8). The maximum amount of bias observed in the data was at θ = +3, where mean bias was −.10; a similar degree of bias was observed at θ = −1.8.

Figure 1
Bias as a Function of θ for Data Sets 1 and 2

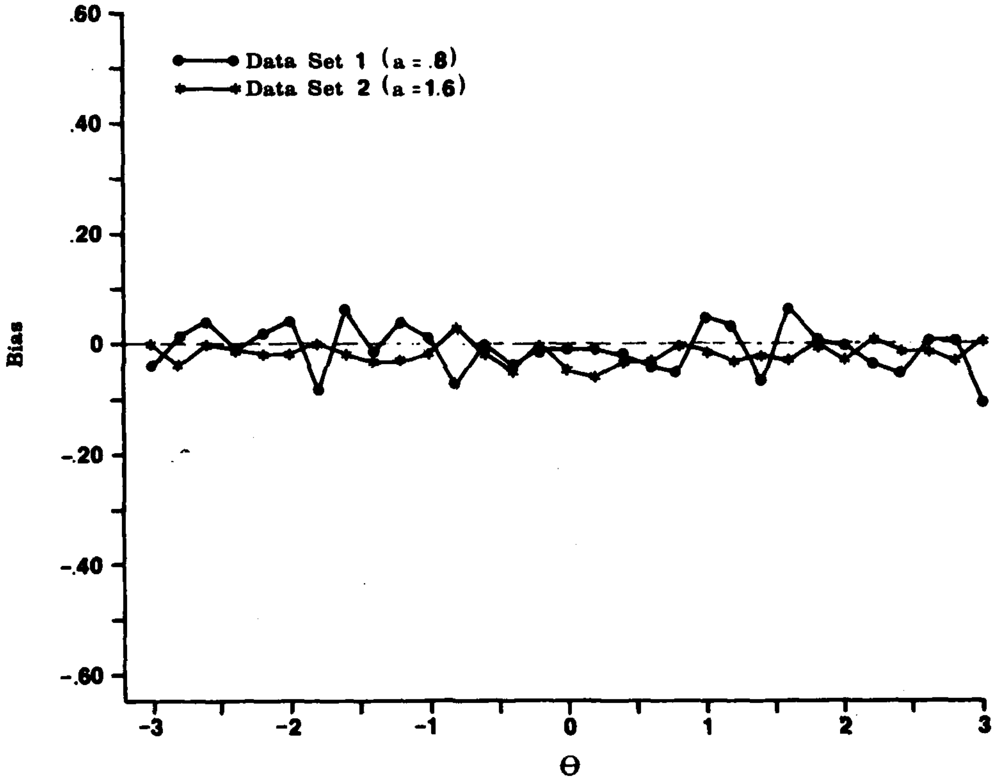
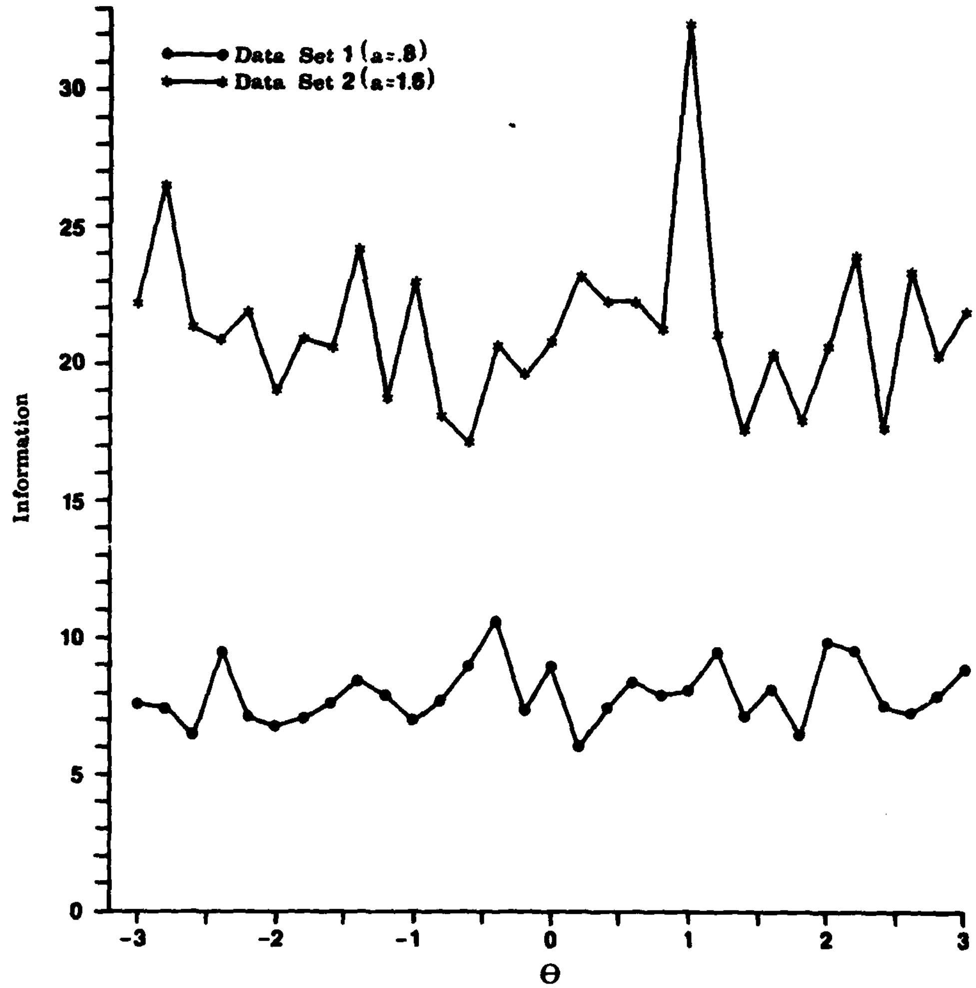


Figure 2 shows information curves for Data Sets 1 and 2. As the results show, the information for Data Set 1 was relatively flat throughout the θ range. The maximum information was observed at θ = −.5, with minimum information at θ = +.2. Information ranged between 7 and 11, with only minor variations across the ability range. The information for Data Set 2 was relatively flat, but not as flat as that for Data Set 1. There was a spike at θ = .8 with a secondary peak at θ = −2.8, and overall more variability between θ levels than for Data Set 1. In general, there is a slight concave trend to the information values for Data Set 2, with the exception of the spike at θ = .8. However, the general trend is a relatively flat information function for both data sets.

Figure 2
Information as a Function of θ for Data Sets 1 and 2



<u>Constant Prior θ Estimate with Fixed Test Length</u>

Figure 3 shows the bias in the θ estimates for the data sets of Study II at each of the three levels of item discrimination (numerical values of bias and information are in Appendix Table B). For all three data sets there is a negative slope to the bias curve with low θ values being overestimated and higher θ values being underestimated. In addition, there are some substantial differences in the bias curves for the three levels of discrimination. Data Set 3 (a = .8) achieved the highest levels of bias of all three data sets. Very severe bias was observed for negative θ levels and severe bias in the opposite direction for positive θ levels. When item discriminations were increased in Data Set 4, there was only a slight drop in the positive bias for low θ levels and a more substantial drop in negative bias for the θ levels above the mean. Increasing the item discriminations to 2.4 in Data Set 5 resulted in virtually no change in bias for low θ level but a further decrease in bias for the positive θ levels with the range of unbiased ability estimates varying from approximately θ

= -1 to θ = +1.5 in Data Set 5. As these results show, the effect of increasing item discrimination is to reduce bias somewhat, primarily for high θ levels. For low θ levels (< -2.0) substantial levels of bias (.20 or more) were observed for the highly discriminating items of Data Set 5.

Figure 3
Bias as a Function of θ for Data Sets 3, 4, and 5

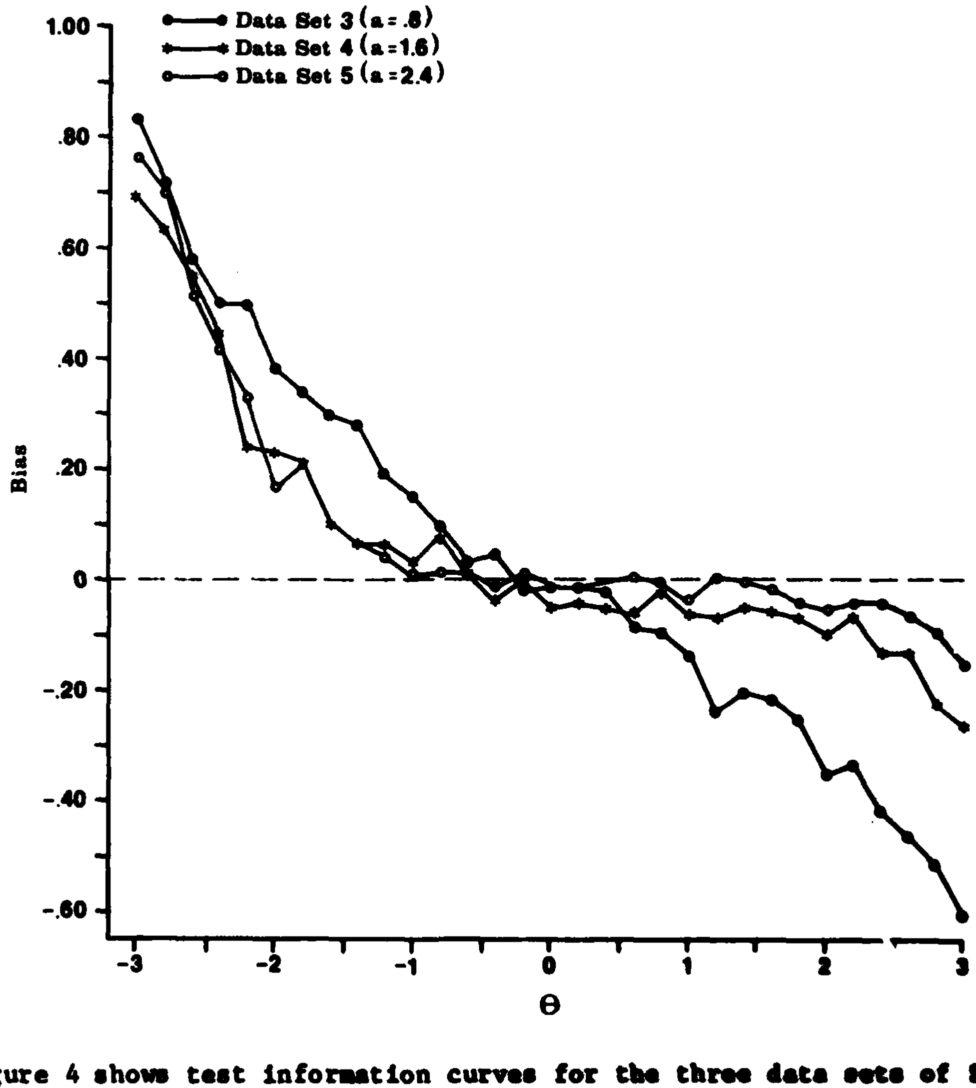
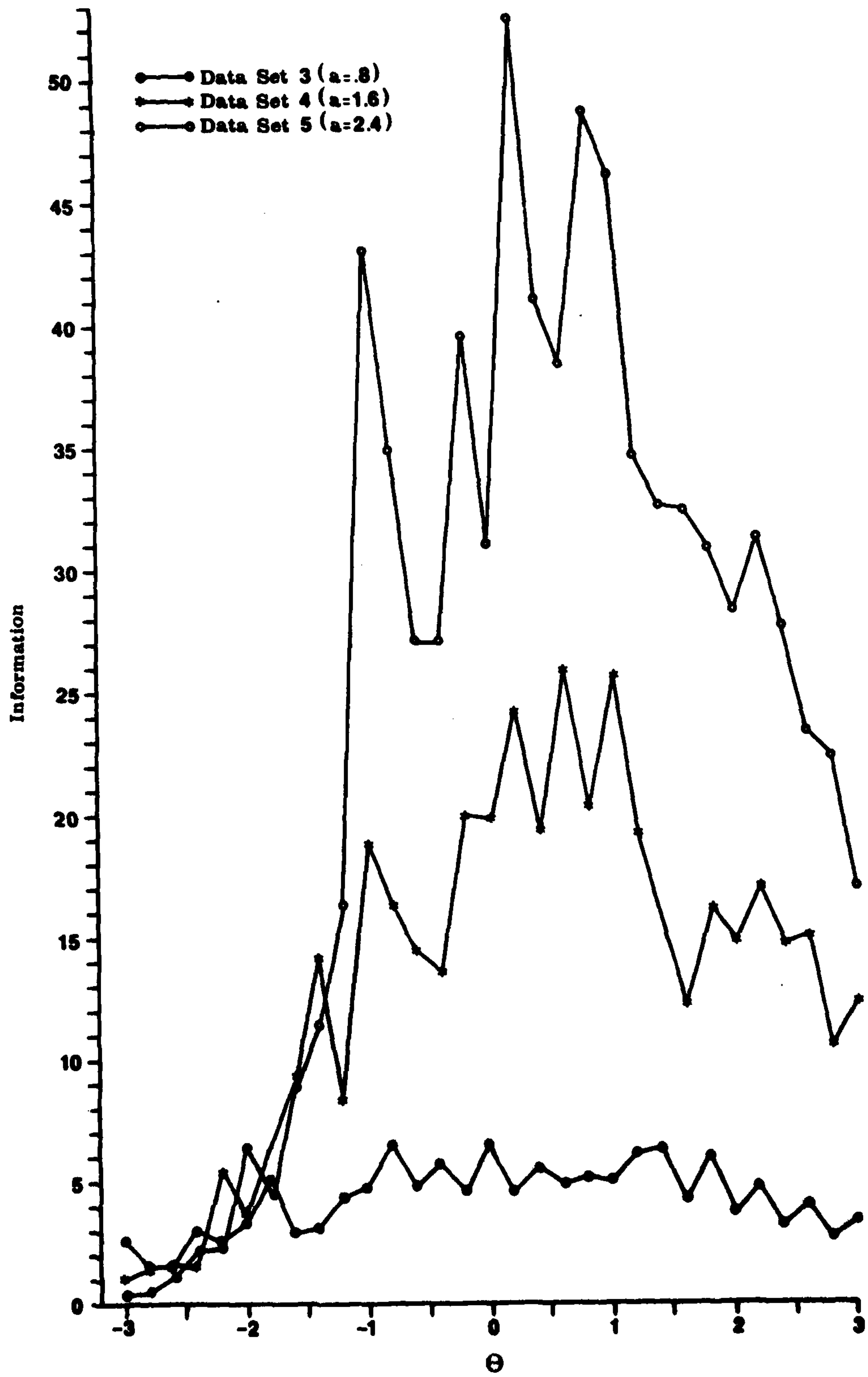


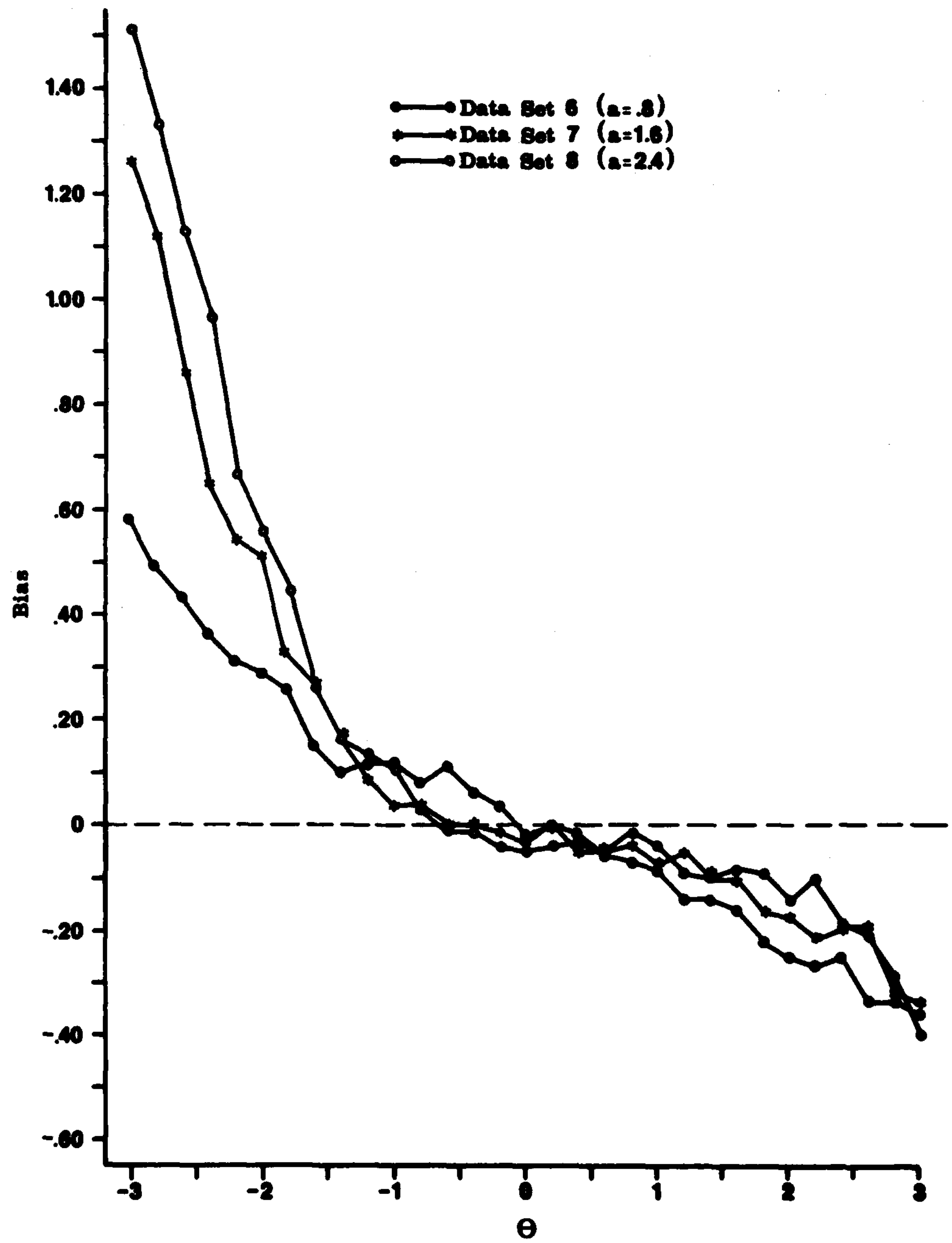
Figure 4 shows test information curves for the three data sets of Study 2. As Figure 4 shows, with the low discriminating items (a = .8) of Data Set 3, test information is relatively flat for θ levels above about θ = -1.5, with a decrease in information below that level. As item discrimination is increased, the results for Data Set 4 show the information curve peaking with relatively lower information levels for θ > 1.6 and θ < -1.5, and a greater asymmetry in the information curve. Finally, when the items of Data Set 5 (a = 2.4) were used, the information curve becomes even more peaked and more variable, with high levels of information generally in the range of θ = +1 to -1, and with information dropping off extremely quickly beyond that range. For θ levels below

Figure 4
Information as a Function of θ for Data Sets 3, 4, and 5



-1, there is little difference in information when item discriminations are increased from a = 1.6 to a = 2.4. For θ levels below -1.8, levels of information are not increased by increasing item discriminations.

Figure 5
Bias as a Function of θ for Data Sets 6, 7, and 8



<u>Constant Prior θ Estimate With Variable Test Length</u>

Figure 5 shows bias functions for the three data sets of Study III (numerical values for bias and information are in Appendix Tables C, D, and E). As the results show, least bias for low θ levels was observed for Data Set 6 (a = .8), while the high θ levels obtained the highest degree of bias for that data set. As item discriminations increased, bias for low θ levels increased, while bias for the high θ levels decreased. Extremely high levels of bias were observed for Data Set 7 (a = 1.6) and Data Set 8 (a = 2.4) for θ levels less than θ = -2.

Figure 6 shows test information functions for the variable-length conditions of Data Sets 6 through 8. The information function that most approximated the horizontal and equiprecise ideal was achieved by Data Set 6 (a = .8), which obtained relatively constant levels of information for θ values greater than θ = -1.5. As item discrimination was increased, the level of information obtained for low θ levels decreased, while the level of information obtained for high θ levels remained similar. The result of increasing item discrimination was a general increase in peakedness and asymmetry of the test information functions.

Figure 6
Information as a Function of θ for Data Sets 6, 7, and 8

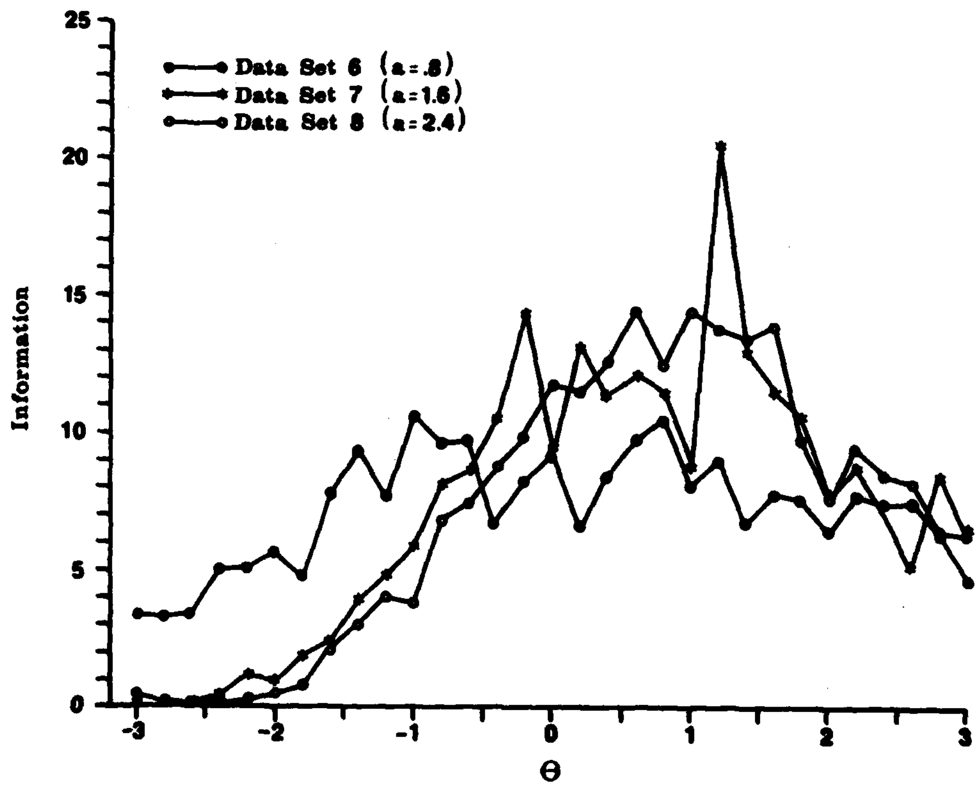
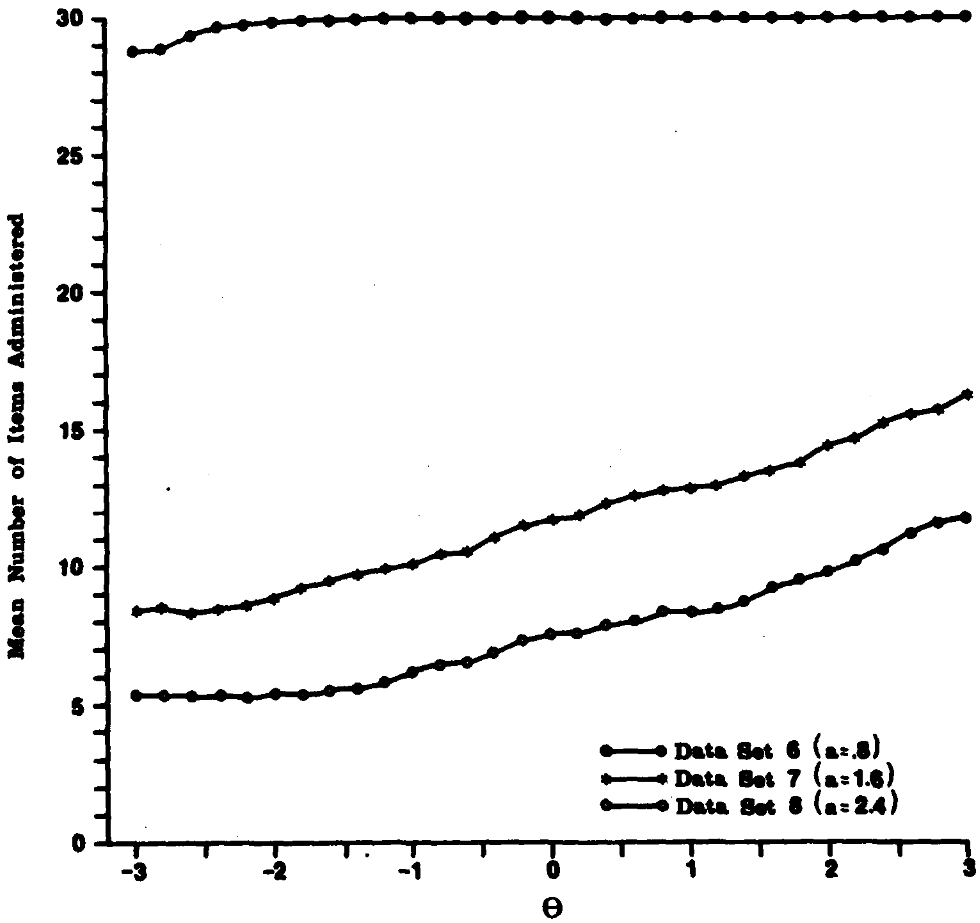


Figure 7 shows the mean number of items administered for each of the θ levels for the data sets of Study III (numerical values are in Appendix Tables C, D, and E). As expected, more items were needed in Data Set 6, which had lower item discriminations, than in Data Sets 7 and 8. The results show that in Data

Set 6, 30 items was generally not sufficient, on the average, for the adaptive test to achieve the specified level of posterior variance (.10) for most test lengths. The results also show that test length required was an increasing function of θ for Data Sets 7 and 8. While, on the average, the posterior variance termination criterion of .10 was achieved with about 8.5 items for low θ values in Data Set 7, twice the number of items (17.0) were necessary to achieve the same posterior variance termination criterion (on the average) for θ = +3. The same trend was observed for the more highly discriminating items of Data Set 8.

Figure 7
Mean Number of Items Administered as a Function of θ
for Data Sets 6, 7, and 8



Discussion and Conclusions

This study used a "perfect" item pool in order to evaluate the performance of Owen's Bayesian adaptive testing strategy under ideal conditions. The results show that in terms of achieving statistically unbiased measurement and measurements of equal precision throughout the range of ability, Owen's adaptive testing strategy achieves these desirable goals only under the extremely unrealistic condition of an accurate prior ability estimate. Of course, in a realistic testing situation, the examinee's ability is not known beforehand; otherwise, testing would not be necessary. Thus, the data of Study 1 serve only as an unrealistic baseline condition to which results of other more realistic testing conditions can be compared. Even under the unrealistic conditions of Study 1, however, there was a tendency for increasing item discrimination to result in increasing variability in levels of information as a function of θ.

Studies II and III evaluated Owen's Bayesian testing strategy under the more realistic testing conditions of a constant prior θ estimate, with both fixed and variable test length. The results of Studies 2 and 3 show that this adaptive testing strategy does not achieve unbiased measurement or measurements of equal precision when a constant prior θ estimate is used for all examinees, regardless of whether test length is fixed or variable. The results show an interaction of the termination criterion with the performance of the adaptive testing strategy, both in terms of bias and information.

When a constant test length is used, increasing item discrimination results in decreased bias, with a more substantial decrease in bias for high θ levels. When variable termination is used, increasing item discrimination results in only slightly decreased bias for high θ levels, but in increased bias for low θ levels, with extremely high levels of bias for very low θ levels. In terms of information, the flattest information curves were observed for both termination criteria with the least discriminating items. As item discrimination was increased, in both cases the information curve became more peaked and asymmetric, with a greater degree of asymmetry observed for the variable-length testing condition. Results also showed that different mean numbers of items were necessary to achieve a fixed posterior variance termination criterion at different levels of θ. With moderately and highly discriminating items (*a* = 1.6 and *a* = 2.4), twice the number of items were necessary, on the average, for high θ levels to reach a posterior variance termination criterion of .10 than for low θ levels.

Because this study used a perfect item pool in which items of a specified discrimination were available at any level of difficulty, the results observed in these studies cannot be attributed to deficiencies in the item pool, as might be the case for the results reported by Gorman (1980). Rather, these results are attributable to the effect of the constant prior θ estimate, as is shown by the comparison of results between Studies II and III and those of Study I. Although the effect of Urry's (1977) correction for regression was not explicitly examined in these studies, it is unlikely that it would have the desired effects under both the fixed-length and variable-length test condition, since, as indicated, there was interaction of observed bias with the termination criterion.

Although a major purpose of adaptive testing is to provide measurements with equal precision/information at all levels of the ability continuum (Weiss, 1982), results of these analyses show that under the realistic conditions of a constant prior θ estimate, Owen's Bayesian adaptive testing strategy does not achieve this desirable goal. Since the test information curves utilize some of the same data from which the bias curves were computed, the results for information are in a sense a consequence of the bias in the θ estimates. The data from these three studies show that the bias results from use of a constant prior θ estimate. Further research will be necessary to determine whether and to what degree the use of variable prior θ estimates will affect the performance of Owen's adaptive testing strategy in terms of reducing the bias and, consequently, improving the equiprecision of its ability estimates.

REFERENCES

Gorman, S. A comparative evaluation of two Bayesian adaptive ability estimation procedures with a conventional test strategy. Unpublished doctoral dissertation, Catholic University of America, Washington DC, 1980.

Jensema, C. J. An application of latent trait mental test theory (Doctoral dissertation, University of Washington, 1972). Dissertation Abstracts International, 1973, 24, 633. (University Microfilms No. 72-20,871)

Jensema, C. J. The validity of Bayesian tailored testing. Educational and Psychological Measurement, 1974, 34, 757-766.

Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. Educational and Psychological Measurement, 1968, 28, 989-1020.

McBride, J. R. Some properties of a Bayesian adaptive ability testing strategy. Applied Psychological Measurement, 1977, 1, 121-140.

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1976.

Owen, R. J. A Bayesian approach to tailored testing (Research Bulletin 69-92). Princeton NJ: Educational Testing Service, 1969.

Owen, R. J. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Urry, V. W. Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement, 1977, 14, 181-196.

Urry, V. W. Individualized testing by Bayesian estimation (Research Bulletin 0171-177). Seattle: University of Washington, Bureau of Testing, April 1971.

Urry, V. W. Computer-assisted testing: The calibration and evaluation of the verbal ability bank (Technical Study 74-3). Washington DC: U.S. Civil Service Commission, Personnel Research and Development Center, December 1974.

Weiss, D. J. Improving measurement quality and efficiency with adaptive testing: Applied Psychological Measurement, 1982, 6, 473-491.

Appendix: Supplementary Tables

Table A
Mean and Variance of $\hat{\theta}$, Bias and Information, as a Function of θ for the Data Sets of Study I

θ	Data Set 1 $\hat{\theta}$ Mean	Data Set 1 $\hat{\theta}$ Variance	Data Set 1 Bias	Data Set 1 Information	Data Set 2 $\hat{\theta}$ Mean	Data Set 2 $\hat{\theta}$ Variance	Data Set 2 Bias	Data Set 2 Information
-3.0	-3.040	.124	-.04	7.669	-3.002	.044	.00	22.253
-2.8	-2.778	.125	.02	7.656	-2.836	.037	-.04	26.509
-2.6	-2.564	.148	.04	6.504	-2.604	.046	.00	21.359
-2.4	-2.406	.102	-.01	9.489	-2.412	.047	-.01	20.939
-2.2	-2.182	.137	.02	7.101	-2.217	.045	-.02	21.905
-2.0	-1.960	.142	.04	6.834	-2.020	.052	-.02	18.985
-1.8	-1.881	.139	-.08	7.061	-1.804	.045	.00	21.972
-1.6	-1.543	.128	.06	7.698	-1.620	.048	-.02	20.629
-1.4	-1.410	.116	-.01	8.523	-1.433	.041	-.03	24.184
-1.2	-1.160	.124	.04	7.934	-1.226	.053	-.03	18.734
-1.0	-.989	.142	.01	7.003	-1.019	.043	-.02	23.121
-.8	-.870	.129	-.07	7.726	-.772	.055	.03	18.099
-.6	-.597	.111	.00	8.996	-.617	.058	-.02	17.184
-.4	-.435	.093	-.04	10.754	-.448	.048	-.05	20.788
-.2	-.208	.135	-.01	7.417	-.197	.051	.00	19.587
0.0	-.010	.110	-.01	9.027	-.052	.048	-.05	20.833
.2	.190	.168	-.01	5.966	.136	.043	-.06	23.279
.4	.379	.133	-.02	7.536	.364	.045	-.03	22.266
.6	.557	.118	-.04	8.491	.570	.045	-.03	22.287
.8	.754	.126	-.05	7.946	.801	.047	.00	21.357
1.0	1.054	.123	.05	8.130	.987	.031	-.01	32.407
1.2	1.226	.105	.03	9.509	1.166	.048	-.03	20.945
1.4	1.333	.141	-.07	7.067	1.379	.057	-.02	17.651
1.6	1.672	.121	.07	8.217	1.570	.049	-.03	20.547
1.8	1.805	.154	.01	6.438	1.796	.056	.00	17.990
2.0	2.003	.108	.00	9.884	1.972	.049	-.03	20.572
2.2	2.168	.103	-.03	9.563	2.213	.042	.01	24.013
2.4	2.353	.128	-.05	7.665	2.390	.057	-.01	17.703
2.6	2.614	.135	.01	7.237	2.585	.043	-.01	23.476
2.8	2.809	.123	.01	7.906	2.774	.050	-.03	20.198
3.0	2.891	.108	-.11	8.958	3.007	.046	.01	21.961

Table B
Mean and Variance of θ, Bias and Information, as a Function of θ
for the Data Sets of Study II

	Data Set 3				Data Set 4				Data Set 5			
	θ			Infor-	θ			Infor-	θ			Infor-
θ	Mean	Variance	Bias	mation	Mean	Variance	Bias	mation	Mean	Variance	Bias	mation
-3.0	-2.166	.103	.83	2.645	-2.308	.161	.69	.945	-2.229	.189	.77	.389
-2.8	-2.084	.193	.72	1.634	-2.169	.162	.63	1.273	-2.097	.228	.70	.544
-2.6	-2.017	.209	.58	1.716	-2.048	.155	.55	1.710	-2.077	.163	.52	1.130
-2.4	-1.896	.133	.50	3.018	-1.957	.215	.44	1.521	-1.992	.114	.41	2.204
-2.2	-1.696	.161	.50	2.755	-1.958	.071	.24	5.505	-1.871	.141	.33	2.296
-2.0	-1.621	.144	.38	3.364	-1.770	.121	.23	3.765	-1.834	.062	.17	6.442
-1.8	-1.463	.103	.34	5.083	-1.582	.080	.22	6.502	-1.588	.104	.21	4.585
-1.6	-1.304	.191	.30	2.936	-1.488	.062	.11	9.410	-1.486	.062	.11	8.940
-1.4	-1.118	.188	.28	3.167	-1.335	.045	.07	14.322	-1.332	.055	.07	11.459
-1.2	-1.008	.143	.19	4.386	-1.128	.084	.07	8.364	-1.147	.043	.05	16.359
-1.0	-.846	.137	.15	4.789	-.972	.040	.03	18.923	-.987	.018	.01	42.925
-.8	-.697	.104	.10	6.554	-.723	.049	.08	16.465	-.781	.024	.02	34.863
-.6	-.567	.146	.03	4.819	-.593	.058	.01	14.682	-.579	.033	.02	27.112
-.4	-.350	.125	.05	5.775	-.432	.065	-.03	13.704	-.414	.035	-.01	27.021
-.2	-.215	.157	-.02	4.689	-.201	.046	.00	20.085	-.193	.025	.01	39.563
0.0	-.014	.115	-.01	6.491	-.032	.048	-.05	19.805	-.009	.033	-.01	31.035
.2	.188	.160	-.01	4.705	.155	.040	-.04	24.265	.192	.020	-.01	52.523
.4	.380	.133	-.02	5.675	.355	.051	-.05	19.288	.404	.026	.00	41.064
.6	.517	.152	-.08	4.952	.544	.038	-.06	26.043	.612	.028	.01	38.412
.8	.715	.143	-.09	5.220	.775	.049	-.02	20.172	.803	.022	.00	48.816
1.0	.866	.147	-.13	5.008	.942	.038	-.06	25.792	.974	.023	-.03	46.216
1.2	.959	.117	-.24	6.169	1.132	.050	-.07	19.294	1.214	.030	.01	34.756
1.4	1.197	.111	-.20	6.339	1.350	.059	-.05	15.974	1.396	.031	.00	32.690
1.6	1.393	.160	-.21	4.260	1.538	.074	-.06	12.345	1.591	.030	-.01	32.517
1.8	1.548	.108	-.25	6.075	1.728	.054	-.07	16.266	1.763	.030	-.04	30.984
2.0	1.650	.174	-.35	3.605	1.898	.056	-.10	14.950	1.951	.031	-.05	28.261
2.2	1.873	.123	-.33	4.840	2.130	.046	-.07	17.189	2.164	.026	-.04	31.384
2.4	1.978	.179	-.42	3.132	2.265	.050	-.13	14.785	2.362	.027	-.04	27.781
2.6	2.144	.130	-.46	4.028	2.466	.045	-.13	15.191	2.538	.029	-.06	23.429
2.8	2.292	.178	-.51	2.721	2.583	.058	-.22	10.766	2.709	.027	-.09	22.413
3.0	2.386	.133	-.61	3.335	2.737	.045	-.26	12.500	2.847	.031	-.15	17.049

Table C
Mean and Variance of $\hat{\theta}$, Bias, Information, and Mean and Standard Deviation of Number of Items Administered as a Function of θ for Data Set 6

θ	$\hat{\theta}$ Mean	$\hat{\theta}$ Variance	Bias	Information	No. of Items Mean	No. of Items S.D.
-3.0	-2.422	.115	.58	3.375	28.67	1.04
-2.8	-2.314	.131	.49	3.281	28.91	1.02
-2.6	-2.166	.138	.43	3.414	29.41	.85
-2.4	-2.038	.101	.36	5.064	29.67	.75
-2.2	-1.894	.109	.31	5.052	29.77	.61
-2.0	-1.707	.103	.29	5.712	29.91	.32
-1.8	-1.543	.131	.26	4.765	29.97	.22
-1.6	-1.450	.084	.15	7.833	29.97	.30
-1.4	-1.297	.073	.10	9.445	29.98	.20
-1.2	-1.076	.093	.12	7.726	30.00	0.00
-1.0	-.876	.069	.12	10.794	30.00	0.00
-.8	-.717	.079	.08	9.723	30.00	0.00
-.6	-.488	.080	.11	9.856	30.00	0.00
-.4	-.338	.117	.06	6.886	30.00	0.00
-.2	-.167	.100	.03	8.195	30.00	0.00
0.0	-.018	.091	-.02	9.120	30.00	0.00
.2	.196	.126	.00	6.642	30.00	0.00
.4	.380	.099	-.02	8.489	30.00	0.00
.6	.540	.086	-.06	9.773	30.00	0.00
.8	.728	.080	-.07	10.462	30.00	0.00
1.0	.922	.103	-.08	8.057	30.00	0.00
1.2	1.055	.090	-.14	9.105	30.00	0.00
1.4	1.261	.119	-.14	6.770	30.00	0.00
1.6	1.438	.100	-.16	7.885	30.00	0.00
1.8	1.578	.101	-.22	7.605	30.00	0.00
2.0	1.749	.118	-.25	6.312	30.00	0.00
2.2	1.929	.092	-.27	7.810	30.00	0.00
2.4	2.149	.093	-.25	7.414	30.00	0.00
2.6	2.271	.087	-.33	7.563	30.00	0.00
2.8	2.466	.100	-.33	6.242	30.00	0.00
3.0	2.639	.124	-.36	4.744	30.00	0.00

Table D
Mean and Variance of $\hat{\theta}$, Bias, Information, and Mean and Standard Deviation of Number of Items Administered as a Function of θ for Data Set 7

θ	$\hat{\theta}$ Mean	$\hat{\theta}$ Variance	Bias	Information	No. of Items Mean	No. of Items S.D.
-3.0	-1.742	.221	1.26	.001	8.37	.90
-2.8	-1.675	.233	1.12	.035	8.49	.85
-2.6	-1.752	.150	.85	.237	8.41	.76
-2.4	-1.762	.152	.64	.523	8.52	.82
-2.2	-1.661	.108	.54	1.263	8.65	.77
-2.0	-1.488	.205	.51	.992	8.96	.86
-1.8	-1.478	.139	.32	1.997	9.30	.91
-1.6	-1.333	.139	.29	2.565	9.45	.75
-1.4	-1.241	.110	.16	3.978	9.85	.77
-1.2	-1.108	.107	.09	4.846	10.03	.77
-1.0	-.955	.103	.04	5.801	10.15	.77
-.8	-.760	.082	.04	8.202	10.62	.81
-.6	-.596	.085	.00	8.731	10.74	.77
-.4	-.402	.077	.00	10.451	11.16	.88
-.2	-.213	.060	-.01	14.320	11.56	.93
0.0	-.028	.099	-.03	9.135	11.81	.96
.2	.195	.071	.00	13.234	11.91	.98
.4	.354	.085	-.05	11.342	12.28	.84
.6	.459	.081	-.05	12.068	12.60	.80
.8	.762	.084	-.04	11.661	12.76	.83
1.0	.930	.110	-.07	8.820	12.91	.88
1.2	1.153	.046	-.05	20.645	12.98	.68
1.4	1.303	.071	-.10	12.934	13.36	.83
1.6	1.504	.076	-.10	11.534	13.65	.91
1.8	1.638	.078	-.16	10.582	13.86	1.00
2.0	1.827	.101	-.17	7.580	14.47	.92
2.2	1.994	.080	-.21	8.730	14.58	.93
2.4	2.210	.089	-.19	7.024	15.13	.82
2.6	2.407	.109	-.19	5.022	15.51	.86
2.8	2.490	.055	-.31	8.490	15.72	.65
3.0	2.675	.063	-.33	6.121	16.17	.87

Table E
Mean and Variance of $\hat{\theta}$, Bias, Information, and Mean and Standard Deviation of Number of Items Administered as a Function of θ for Data Set 8

θ	$\hat{\theta}$ Mean	$\hat{\theta}$ Variance	Bias	Information	No. of Items Mean	No. of Items S.D.
-3.0	-1.485	.216	1.51	.417	5.33	.57
-2.8	-1.473	.230	1.33	.117	5.31	.54
-2.6	-1.466	.183	1.13	.007	5.29	.55
-2.4	-1.432	.284	.97	.026	5.31	.54
-2.2	-1.528	.178	.67	.222	5.22	.50
-2.0	-1.439	.185	.56	.503	5.55	.58
-1.8	-1.354	.193	.45	.844	5.44	.59
-1.6	-1.345	.113	.26	2.168	5.50	.56
-1.4	-1.227	.113	.17	2.964	5.67	.55
-1.2	-1.056	.108	.14	3.973	5.91	.45
-1.0	-.886	.139	.11	3.771	6.15	.62
-.8	-.768	.091	.03	6.780	6.39	.69
-.6	-.615	.095	-.01	7.419	6.50	.75
-.4	-.409	.090	-.01	8.725	6.95	.86
-.2	-.240	.087	-.04	9.841	7.28	.78
0.0	-.048	.078	-.05	11.742	7.43	.67
.2	.157	.084	-.04	11.463	7.61	.61
.4	.368	.079	-.03	12.611	7.93	.65
.6	.548	.070	-.05	14.501	8.01	.68
.8	.794	.082	-.01	12.427	8.27	.83
1.0	.956	.070	-.04	14.400	8.25	.73
1.2	1.111	.071	-.09	13.834	8.48	.77
1.4	1.299	.071	-.10	13.272	8.78	.88
1.6	1.519	.064	-.08	13.892	9.23	.86
1.8	1.708	.085	-.09	9.693	9.56	.72
2.0	1.859	.100	-.14	7.482	9.83	.72
2.2	2.099	.071	-.10	9.353	10.26	.74
2.4	2.224	.069	-.18	8.312	10.61	.82
2.6	2.393	.059	-.21	8.124	11.10	.89
2.8	2.517	.060	-.28	6.404	11.44	.80
3.0	2.605	.047	-.39	6.204	11.75	.61

Distribution List

Navy

1 Liaison Scientist
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510

1 Lt. Alexander Bory
Applied Psychology
Measurement Division
NAMRL
NAS Pensacola, FL 32508

1 Dr. Stanley Collyer
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 Mike Durmeyer
Instructional Program Development
Building 90
NET-PDCD
Great Lakes NTC, IL 60088

1 DR. PAT FEDERICO
Code P13
NPRDC
San Diego, CA 92152

1 Dr. Cathy Fernandes
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. William L. Maloy (02)
Chief of Naval Education and Training
Naval Air Station

Pensacola, FL 32508

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Cdr Ralph McCumber
Director, Research & Analysis Division
Navy Recruiting Command
4015 Wilson Boulevard
Arlington, VA 22203

1 Dr. George Moeller
Director, Behavioral Sciences Dept.
Naval Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 63409

1 Dr William Montague
NPRDC Code 13
San Diego, CA 92152

1 Bill Nordbrock
1032 Fairlawn Ave.
Libertyville, IL 60048

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychological Sciences Division
Code 442
Office of Naval Research
Arlington, VA 22217

6 Personnel & Training Research Group
Code 442PT
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Office of the Chief of Naval Operations
Research Development & Studies Branch
OP 115
Washington, DC 20350

1 LT Frank C. Petho, MSC, USN (Ph.D)
CNET (N-432)
NAS
Pensacola, FL 32508

1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940

1 Dr. Bernard Rimland (01C)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

1 Dr. Worth Scanland
CNET (N-5)
NAS, Pensacola, FL 32508

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frederick Steinheiser
CNO - OP115
Navy Annex
Arlington, VA 20370

1 Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frank Vicino
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Edward Wegman
Office of Naval Research (Code 411S&P)
800 North Quincy Street
Arlington, VA 22217

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

Marine Corps

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower Utilizatio
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 Headquarters, U. S. Marine Corps
Code MPI-20
Washington, DC 20380

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

1 Major Frank Yohannan, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

Army

1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Director, Training Research Lab
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joyce Shields
Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Hilda Wing
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

1 Dr. Robert Wisher
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 AFHRL/LRS
Attn: Susan Ewing
WPAFB
WPAFB, OH 45433

1 Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

1 U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

1 Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

1 Dr. Roger Pennell
Air Force Human Resources Laboratory
Lowry AFB, CO 80230

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Department of Defense

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

1 Jerry Lehnus
HQ MEPCOM
Attn: MEPCT-P
Fort Sheridan, IL 60037

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defens
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Dr. Wayne Sellman
Office of the Assistant Secretary
of Defense (MRA & L)
2B269 The Pentagon
Washington, DC 20301

Civilian Agencies

1 Dr. Helen J. Christup
Office of Personnel R&D
1900 E St., NW
Office of Personnel Management
Washington, DC 20015

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Chief, Psychological Research Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Private Sector

1 Dr. James Algina
University of Florida
Gainesville, FL 326

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 Psychological Research Unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600
AUSTRALIA

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Capt. J. Jean Belanger
Training Development Division
Canadian Forces Training System
CFTSHQ, CFB Trenton
Astra, Ontario, KOK
CANADA

1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
Israel

1 Dr. Werner Birke
DezWPs im Streitkraefteamt
Postfach 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. R. Darrel Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Mr. Arnold Bohrer
Section of Psychological Research
Caserne Petits Chateau
CRS
1000 Brussels
Belgium

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52243

1 Bundministerium der Verteidigung
-Referat P II 4-
Psychological Service
Postfach 1328
D-5300 Bonn 1
F. R. of Germany

1 Dr. Ernest R. Cadotte
307 Stokely
University of Tennessee
Knoxville, TN 37916

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. Hans Crombag
Education Research Center
University of Leyden
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

1 Dr. Kenneth B. Cross
Anacapa Sciences, Inc.
P.O. Drawer Q
Santa Barbara, CA 93102

1 Dr. Walter Cunningham
University of Miami
Department of Psychology
Gainesville, FL 32611

1 Dr. Dattpradad Divgi
Syracuse University
Department of Psychology
Syracuse, NE 33210

1 Dr. Fritz Drasgow
Department of Psychology
University of Illinois
603 E. Daniel St.
Champaign, IL 61820

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Benjamin A. Fairbank, Jr.
McFann-Gray & Associates, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 Dr. Dexter Fletcher
WICAT Research Institute
1875 S. State St.
Orem, UT 22333

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01002

1 Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
3939 O'Hara Street
PITTSBURGH, PA 15260

1 Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Delwyn Harnisch
University of Illinois
242b Education
Urbana, IL 61801

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Jack Hunter
2122 Coolidge St.
Lansing, MI 48906

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Dr. Douglas H. Jones
Room T-255/21-T
Educational Testing Service
Princeton, NJ 08541

1 Professor John A. Keats
University of Newcastle
N. S. W. 2308
AUSTRALIA

1 Dr. Scott Kelso
Haskins Laboratories, Inc
270 Crown Street
New Haven, CT 06510

1 CDR Robert S. Kennedy
Canyon Research Group
1040 Woodcock Road
Suite 227
Orlando, FL 32803

1 Dr. William Koch
University of Texas-Austin
Measurement and Evaluation Center
Austin, TX 78703

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
Netherlands

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Mr. Phillip Livingston
Systems and Applied Sciences Corporatio
6811 Kenilworth Avenue
Riverdale, MD 20840

1 Dr. Robert Lockman
Center for Naval Analysis
200 North Beauregard St.
Alexandria, VA 22311

1 Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

1 Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Mr. Robert McKinley
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

1

Professor Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Robert Mislevy
711 Illinois Street
Geneva, IL 60134

1 Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

1 Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

1 Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Cirle, NW
Washington, DC 20036

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 Mr. L. Petrullo
3695 N. Nelson St.
ARLINGTON, VA 22207

1 Dr. Richard A. Pollak
Director, Special Projects
Minnesota Educational Computing
2520 Broadway Drive
St. Paul,MN

1 Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

1 Dr. Thomas Reynolds
University of Texas-Dallas
Marketing Department
P. O. Box 688
Richardson, TX 75080

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 Frank L. Schmidt
Department of Psychology
Bldg. GG
George Washington University
Washington, DC 20052

1 Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. William Sims
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

1 Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Lab
252 Engineering Research Laboratory
Urbana, IL 61801

1 Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. V. R. R. Uppuluri
Union Carbide Corporation
Nuclear Division
P. O. Box Y
Oak Ridge, TN 37830

1 Dr. David Vale
Assessment Systems Corporation
2233 University Avenue
Suite 310
St. Paul, MN 55114

1 Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08540

1 Dr. Michael T. Waller
Department of Educational Psychology
University of Wisconsin--Milwaukee
Milwaukee, WI 53201

1 Dr. Brian Waters
HumRRO
300 North Washington
Alexandria, VA 22314

1 DR. GERSHON WELTMAN
PERCEPTRONICS INC.
6271 VARIEL AVE.
WOODLAND HILLS, CA 91367

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
Lawrence, KS 66045

1 Dr. Rand R. Wilcox
University of Southern California
Department of Psychology
Los Angeles, CA 90007

1 Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. Bruce Williams
Department of Educational Psychology
University of Illinois
Urbana, IL 61801

1 Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

PREVIOUS PUBLICATIONS (CONTINUED)

78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978.
77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977.
77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977.
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977.
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977.
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977.
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977.
Final Report: Computerized Ability Testing, 1972-1975. April 1976.
76-5. Effects of Item Characteristics on Test Fairness. December 1976.
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976.
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976.
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976.
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975.
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975.
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975.
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975.
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975.
74-5. Strategies of Adaptive Ability Measurement. December 1974.
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974.
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974.
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974.
73-4. An Empirical Study of Computer-Administered Two-Stage Ability Testing. October 1973.
73-3. The Stratified Adaptive Computerized Ability Test. September 1973.
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973.

Copies of these reports are available, while supplies last, from:
Computerized Adaptive Testing Laboratory
N660 Elliott Hall
University of Minnesota
75 East River Road
Minneapolis MN 55455 U.S.A.

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

81-1. Reliability and Validity of Adaptive and Conventional Tests in a Military Recruit Population. January 1983.
81-5. Dimensionality of Measured Achievement Over Time. December 1981.
81-4. Factors Influencing the Psychometric Characteristics of an Adaptive Testing Strategy for Test Batteries. November 1981.
81-3. A Validity Comparison of Adaptive and Conventional Strategies for Mastery Testing. September 1981.
Final Report: Computerized Adaptive Ability Testing. April 1981.
81-2. Effects of Immediate Feedback and Pacing of Item Presentation on Ability Test Performance and Psychological Reactions to Testing. February 1981.
81-1. Review of Test Theory and Methods. January 1981.
80-5. An Alternate-Forms Reliability and Concurrent Validity Comparison of Bayesian Adaptive and Conventional Ability Tests. December 1980.
80-4. A Comparison of Adaptive, Sequential, and Conventional Testing Strategies for Mastery Decisions. November 1980.
80-3. Criterion-Related Validity of Adaptive Testing Strategies. June 1980.
80-2. Interactive Computer Administration of a Spatial Reasoning Test. April 1980.
Final Report: Computerized Adaptive Performance Evaluation. February 1980.
80-1. Effects of Immediate Knowledge of Results on Achievement Test Performance and Test Dimensionality. January 1980.
79-7. The Person Response Curve: Fit of Individuals to Item Characteristic Curve Models. December 1979.
79-6. Efficiency of an Adaptive Inter-Subtest Branching Strategy in the Measurement of Classroom Achievement. November 1979.
79-5. An Adaptive Testing Strategy for Mastery Decisions. September 1979.
79-4. Effect of Point-in-Time in Instruction on the Measurement of Achievement. August 1979.
79-3. Relationships among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods. April 1979.
Final Report: Bias-Free Computerized Testing. March 1979.
79-2. Effects of Computerized Adaptive Testing on Black and White Students. March 1979.
79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979.
78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978.
78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.
78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978.
78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.

-continued inside-